TESTING FOR PERIODICALLY COLLAPSING BUBBLES: AN GENERALIZED SUP ADF TEST

SHU-PING SHI

School of Economics, The Australian National University

SUMMARY

Identifying explosive bubbles under the influence of their periodically collapsing property has long been a concern in bubble testing literature. In this paper, we argue that the sup Augmented Dickey-Fuller (ADF) test (Phillips, Wu and Yu, 2009), which implements a right-tail ADF test and a sup test on a forward expanding sample sequence, is sensitive to the sample starting point when there are more than one bubble collapsing episodes within the sample range. To surmount this pitfall we propose an alternative method named the generalized sup ADF test, which amplifies the sample sequence by varying the sample starting point within its feasible range. This test improves the power of the bubble testing method significantly. We then apply both tests to the Hong Kong stock market from October 1980 to April 2009. The generalized sup ADF tests find evidence of explosive behavior in the Hang Seng Index, whereas the sup ADF tests suggest the opposite.

Key words: rational bubble, periodically collapsing, sup ADF test, generalized sup ADF test

JEL classification: C22

Email address: shuping.shi@anu.edu.au (SHU-PING SHI).

[1] I am very grateful to Jun Yu and Peter C.B. Phillips for advice in the early stage of this research at Singapore Management University. I thank Heather Anderson, Farshid Vahid and Tom Smith for many valuable discussions and suggestions.

21 January 2010

1 INTRODUCTION

The literature on the identification of rational bubbles from market fundamentals stems from the Lucas asset pricing model. Most econometric tests of bubbles, namely, the cointegration based test (Diba and Grossman, 1988), West's two-step test (West, 1987), the variance bounds test (Shiller, 1981, LeRoy and Porter, 1981) and the intrinsic bubbles test (Froot and Obsterfeld, 1991), begin with the following equation (for an overview of econometric tests of bubbles, see Gurkaynak (2008)):

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f} \right)^i \mathbb{E}_t \left(D_{t+i} \right) + B_t \tag{1}$$

where P_t is the after-dividend price of the asset (i.e stock price), and D_t is the payoff received from the asset (i.e. dividend) and r_f is the risk-free interest rate. B_t defines the bubble component, which has an explosive property

$$\mathbb{E}_t \left(B_{t+1} \right) = \left(1 + r_f \right) B_t. \tag{2}$$

This equation implies that bubbles cannot pop and restart (Diba and Grossman, 1998). Provided that negative asset prices are impossible (Tirole, 1982; Wu, 1997), a multiplicative form between B_t and ε_t is more reasonable than an additive form. That is, $B_{t+1} = \left(1 + r_f \right) B_t \varepsilon_{t+1}$, where $\mathbb{E} \left(\varepsilon_t \right) = 1$. Therefore, if B_t equals zero at time t, it will stay at zero for all future periods.

However, Evans (1991) argues that it is possible that bubbles collapse to a non-zero value and continue to grow at some explosive rate depending on the bubble size. [2] Furthermore, Evans (1991) shows via simulation that the conventional cointegration based test, which relies on a right-tail unit root test (with an explosive alternative hypothesis), is incapable of detecting explosive bubbles under the influence of the periodically collapsing property. [3]

[2] Blanchard (1979) also notes the periodically collapsing property of bubbles.
[3] The failure of the conventional cointegration based test is further studied in Charemza and Deadman (1995) with the setting of bubbles with stochastic explosive roots.

This argument has led to a number of papers which propose bubble testing methods that have some power in detecting periodically collapsing bubbles. One of the prevalent methods is the sup ADF test (or the forward recursive ADF test) put forward by Phillips, Wu and Yu (2009, PWY hereafter). They propose to implement the unit root test repeatedly on a forward expanding sample sequence and make inference based on the sup value of the corresponding ADF statistic sequence. They show that, compared to the conventional stationarity test, the sup ADF test improves the power significantly in the presence of periodically collapsing bubbles.

In this paper, we argue that the testing value of the sup ADF test relies greatly on the starting points of samples. Namely, if the starting point of a sample is selected so that the sample includes more than one bubble collapsing episodes, the test may fail to reveal the existence of bubbles.

To overcome this pitfall of the sup ADF test, we propose an alternative method named *the generalized sup ADF test*. The generalized sup ADF test is also based on the idea of repeatedly implementing the ADF test; however, it extends the sample sequence to a broader range. Instead of fixing the starting points of the samples (namely, on the first observation of the total sample), the generalized sup ADF test extends the sample sequence by changing the starting point of each sample over a feasible range, and superimposing expanding sample sequences onto each starting point. Consistent with the sup ADF test, the sample sequence is designed (i) to capture the explosive phase within the total sample and (ii) to ensure that there are sufficient observations to achieve estimation efficiency. Therefore, the generalized sup ADF test, which covers more samples, is expected to outperform the sup ADF test in finding the most explosive phase with the total sample, given an identical smallest sample size. The asymptotic distribution of the generalized sup ADF statistic is then compared to that of the sup ADF test.

The improvement of the generalized sup ADF test over the sup ADF test is demonstrated by performing both tests on a simulated asset price series. Furthermore, based on the Lucas asset pricing model and the Evans' bubble model, we calculate the powers of these two methods and find a significant gain in the generalized sup ADF test. We then apply the sup ADF test and the generalized sup ADF test to the Hong Kong stock market from October 1980 to April 2009.

The outline of this chapter is as follows. Section 2 discusses the rationale of the conventional cointegration based bubble test. The sup ADF test and the generalized sup ADF test, along with the asymptotic distribution of the sup ADF statistic and the generalized sup ADF statistic, are described in Section 3. Section 4 explores the sensitivity of the sup ADF test to the starting point of the estimation sample via experimenting on simulated asset prices. We then implement the generalized sup ADF test on the same simulated data series to show the advantage of the test. Power comparison is conducted in Section 5. An application of these tests to the Hong Kong Stock index is presented in Section 6. Section 7 concludes the paper.

2 THE CONVENTIONAL COINTEGRATION BASED TEST

Based on the explosive property of bubbles, Diba and Grossman (1988) recommend the strategy of using a stationarity test for the logarithmic asset prices and observable market fundamentals, such as the logarithmic dividends. The conventional stationarity test is based on the standard Augmented-Dickey-Fuller test or Phillips-Perron test (Phillips and Perron, 1998), but has an explosive alternative hypothesis. Consider the model

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{k} \psi_i \Delta y_{t-i} + \varepsilon_t \tag{3}$$

where y_t is the logarithmic asset price or the logarithmic dividend, $\varepsilon_t \sim N(0, \sigma^2)$ and k is the number of lags. The significance test (Ng and Perron, 2001) is used to determine the lag order. The null hypothesis is $\beta = 0$, which implies that y_t is a unit root process (Δy_t is stationary). The alternative hypothesis is $\beta > 0$, meaning that y_t is explosive (Δy_t is non-stationary).

When there are no bubbles in the market, equation (1) implies that

$$\left(1-\rho\right)p_t^f-\rho e^{\bar{d}-\bar{p}}d_t=\kappa+e^{\bar{d}-\bar{p}}\sum_{j=1}^{\infty}\rho^j\mathbb{E}_t\left[\triangle d_{t+j}\right],\ \text{(see Appendix A.1)} \tag{4}$$

where $p_t = \log(P_t)$, $d_t = \log(D_t)$, $\rho = (1+r_f)^{-1}$ and $\kappa = (\rho-1)(1-\bar{p}) + \rho e^{\bar{d}-\bar{p}}\left(1-\bar{d}\right)$, where $\bar{p}$ and $\bar{d}$ are the respective sample means of p_t and d_t. This equation manifests the rationale of the cointegration based bubble tests. If the first order difference of the logarithmic dividend $\triangle d_t$ is stationary, p_t and d_t should be cointegrated with vector $[(1-\rho), -\rho e^{\bar{d}-\bar{p}}]$ in normal market states.

Due to the possible presence of unobservable market fundamentals such as intangible capital (Li, 2005), we cannot use evidence of nonstationarity in the first order difference of the asset prices $\triangle p_t$ to conclude that there are bubbles. However, the reverse inference can be established. Namely, if no evidence of nonstationarity is found, the possibility of bubbles can be ruled out.

3 THE SUP ADF TEST AND THE GENERALIZED SUP ADF TEST

Suppose the regression sample starts from the r_1^{th} fraction of the total sample and ends at fraction r_2, where $r_2 = r_1 + r_w$ and r_w is the fraction of the sample size in the regression. The number of observations in the regression is $T_w = [Tr_w]$, where $[.]$ signifies the integer part of its argument and T is the total number of observations.

The sup ADF test proposed by PWY implements the ADF test repeatedly on a forward expanding sample sequence. The starting point r_1 of the sample sequence is fixed at 0, so the ending point of each sample r_2 is equal to r_w. The sample window r_w expands from r_0 to 1, where r_0 is the smallest sample window (selected to ensure estimation efficiency) and 1 is the largest sample window (total sample size). The sup ADF statistic is defined as $\sup_{r_w\in[r_0,1]} ADF_{r_w}$, and it is denoted by $SADF$. Under the null hypothesis that the true process is a random walk without drift, the asymptotic distribution of the sup ADF statistic is

$$SADF \xrightarrow{L} \sup_{r_w \in [r_0,1]} \left\{ \frac{r_w \left[\int_0^{r_w} W dW - \frac{1}{2} r_w \right] - \int_0^{r_w} W dr . W(r_w)}{r_w^{1/2} \left\{ r_w \int_0^{r_w} W^2 dr - \left[\int_0^{r_w} W(r) dr \right]^2 \right\}^{1/2}} \right\},$$

where W is the standard Wiener process (see appendix B and C.1 for the proof).

Compared with the sup ADF test, the generalized sup ADF test extends the sample sequence to include more samples. Besides expanding the sample window r_w, the generalized sup ADF test allows the sample starting point r_1 to vary within its feasible range, which is from 0 to $1 - r_w$. The regression starts from the first observation when $r_1 = 0$, and when $r_1 = 1 - r_w$, the regression sample covers the last observation. The respective ADF statistic is denoted by $ADF_{r_1}^{r_w}$. We define the generalized sup ADF statistic to be the largest ADF statistic over the feasible ranges of r_w and r_1, and we denote this statistic by $GSADF$. That is,

$$GSADF = \sup_{r_w \in [r_0,1]} \left\{ \sup_{r_1 \in [0, 1-r_w]} ADF_{r_1}^{r_w} \right\},$$

and the corresponding window size of the $GSADF$ is referred to as the optimum window size r_w^*, namely

$$r_w^* = \arg \sup_{r_w \in [r_0,1]} \left\{ \sup_{r_1 \in [0, 1-r_w]} ADF_{r_1}^{r_w} \right\}.$$

Under the null hypothesis that the true process is a random walk without drift, the asymptotic distribution of the generalized sup ADF statistic is (see appendix C.1)

$$\sup_{r_w \in [r_0,1]} \sup_{\substack{r_1 \in [0,1-r_w] \\ r_2 = r_1 + r_w}} \left\{ \frac{r_w \left[\int_{r_1}^{r_2} W dW - \frac{1}{2} r_w \right] - \int_{r_1}^{r_2} W(r) dr . \left[W(r_2) - W(r_1) \right]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W^2 dr - \left[\int_{r_1}^{r_2} W dr \right]^2 \right\}^{1/2}} \right\}.$$

It is well known that the Wiener process has independent increments with distribution $W(r_2) - W(r_1) \sim N(0, r_w)$. We can then infer that the generalized sup ADF test nests the sup ADF test.

Suppose the true process is a random walk with drift, then both the sup ADF statistic and the generalized sup ADF converge to the standard normal distribution. Thus, $SADF$ and $GSADF$ test statistics can be compared to the usual t tables to perform an asymptotically valid test (see appendix C.2 for the proof).

In practice, r_0 is inversely related to the total number of observations T. If T is small, r_0 needs to be large enough to achieve estimation efficiency. If T is large, r_0 can be set to be a smaller number so that we will not miss any opportunity to capture the most explosive phase.

To obtain the asymptotic critical values of the ADF statistic distributions under the null hypothesis that the true process is a random walk, we resort to simulation. One of the key steps is to simulate the standard Wiener process. Since the Wiener process is continuous and stochastic, we can only generate a path sampled with a finite number of points. Suppose that $t_1, t_2, \cdots, t_N$ are equally spaced within a finite interval. At each point, we generate a Gaussian random variable with mean 0 and variance $1/N$. The value of $W(r)$ is the sum of the first r increments.

The asymptotic critical values under the null hypothesis that the true process is a random walk without drift are displayed in Table 1. The simulated asymptotic

Table 1. The asymptotic critical values of the ADF tests with constant (true process is a random walk without drift)

	ADF	SADF	GSADF
	Stationary Alternative $H_1 : \beta < 0$		
1%	-3.47	-2.64	-4.98
5%	-2.89	-2.11	-2.80
10%	-2.59	-1.70	-1.73
	Explosive Alternative $H_1 : \beta > 0$		
10%	-0.46	0.51	7.39
5%	-0.11	0.81	10.30
1%	0.56	1.32	13.16

Note: The number of discrete points for approximating the Wiener process and the integral are 5,000 and 2,000 respectively. The smallest sample fraction r_0 for the sup ADF statistic and the generalized sup ADF statistic is 0.1.

critical values for the ADF test are consistent with those in Fuller (1996, Table 10.A.2). The right-tail critical values of the generalized sup ADF test are larger than those of the sup ADF test.

4 SIMULATION STUDY

In this section, we demonstrate how the the sup ADF test and the generalized sup ADF test work when we testing a sample period that contains more than one bubble collapses.

4.1 Generating the test sample

We first simulate an asset price series based on the Lucas asset pricing model and the Evans's bubble model. The simulated asset prices consist of a market fundamental component P_t^f, which combines a random walk dividend process and the Lucas asset pricing equation[4] to obtain (see Appendix A.2)

$$D_t = \mu + D_{t-1} + \varepsilon_{Dt}, \varepsilon_{Dt} \sim N\left(0, \sigma_D^2\right) \tag{5}$$

$$P_t^f = \frac{\mu\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho} D_t \tag{6}$$

and a bubble component proposed in Evans (1991) such that

$$B_{t+1} = \rho^{-1} B_t \varepsilon_{B,t+1}, \quad \text{if} \;\; B_t < b \tag{7}$$

$$B_{t+1} = \left[\zeta + (\pi\rho)^{-1} \theta_{t+1} \left(B_t - \rho\zeta\right)\right] \varepsilon_{B,t+1}, \;\; \text{if} \;\; B_t \geq b \cdot \tag{8}$$

[4] An alternative data generating process, which assumes that the logarithmic dividend is a random walk with drift, is as follows:

$$\ln D_t = \mu + \ln D_{t-1} + \varepsilon_t, \varepsilon_t \sim N\left(0, \sigma_d^2\right)$$
$$P_t^f = \frac{\rho \exp\left(\mu + \frac{1}{2}\sigma_d^2\right)}{1 - \rho \exp\left(\mu + \frac{1}{2}\sigma_d^2\right)} D_t.$$

This has the property that $\mathbb{E}_t(B_{t+1}) = (1+r_f) B_t$. μ is the drift of the dividend process, σ_D^2 is the variance of the dividend, $\rho^{-1} = 1 + r_f > 1$ and $\varepsilon_{B,t} = \exp(y_t - \tau^2/2)$ with $y_t \sim NID(0, \tau^2)$. ζ is the remaining size after the bubble collapse. θ_t follows a Bernoulli process which takes the value 1 with probability π and 0 with probability $1-\pi$. That is, θ_t takes the value 1 if the bubbles survive at period t; otherwise, it takes 0. Equation (7) states that a bubble grows explosively at rate ρ^{-1} when its size is less than b. If the size is greater than b (equation (8)), the bubble grows at a faster rate $((\pi\rho)^{-1} > \rho^{-1})$ but with $1-\pi$ probability of collapsing.

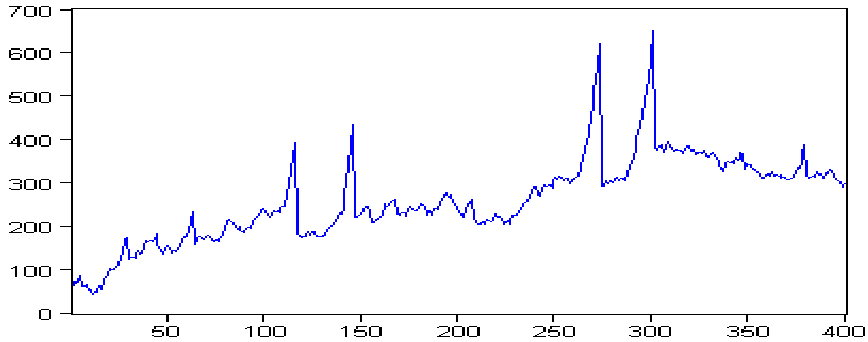


Fig. 1. Simulated data with sample size=400

We set the parameters in the data generating process as in Evans (1991): $B_0 = 0.5, \alpha = 1, \pi = 0.85, \zeta = 0.5, \rho = 0.952, \tau = 0.05, \mu = 0.0373, D_0 = 1.3$, and $\sigma_D^2 = 0.1574$. Stock prices are then calculated as the sum of the market fundamental component and the bubble component: $P_t = P_t^f + 20B_t$ as in Evans (1991). The sample size T is set to 400. Figure 1 depicts one realization of the data generating process. As we can observe from this graph, there are four obvious spires within the sample. Those spires are either results of bubbles collapsing or bubble-like volatilities in asset prices.

4.2 Generating the appropriate critical values

Since the data generating process involves a constant term, we simulate critical values for these tests under the null hypothesis that the true process is a random walk with drift. Critical values displayed in Table 2 are obtained from 5,000 Monte

Carlo simulations with 400 observations. The maximum lag order of the significant test is set to 12 for all tests in this paper. The smallest window size r_0 considered is 0.1, which contains 40 observations. As we can see from the table, the 5% right-tail critical value of the ADF test, the sup ADF test and the generalized sup ADF test are 1.26, 2.85 and 4.75 respectively.

Table 2. Critical values of the ADF tests with constant (true process is a random walk with drift)

	ADF	SADF	GSADF
	Stationary Alternative $H_1 : \kappa < 0$		
1%	-2.92	-0.45	1.67
5%	-2.16	0.07	2.11
10%	-1.75	0.33	2.34
	Explosive Alternative $H_1 : \kappa > 0$		
10%	0.88	2.49	4.39
5%	1.26	2.85	4.75
1%	1.99	3.54	5.52

Note: critical values of both tests are obtained from 5,000 Monte Carlo simulations with sample size 400. The maximum lag order is set to 12. The smallest sample has 40 observations ($r_0 = 0.1$).

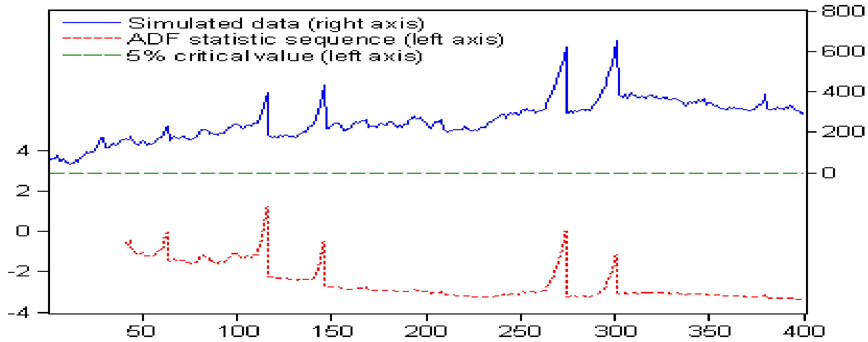
4.3 Performing the sup ADF test and the generalized sup ADF test

In this subsection, we first implement the sup ADF test on the whole sample range. To illustrate the instability of the sup ADF test, we repeat the test on a sub-sample which contains fewer spires. Furthermore, to show the advantage of the generalized sup ADF test, we conduct the test on the same simulated data series (with the whole sample rang).

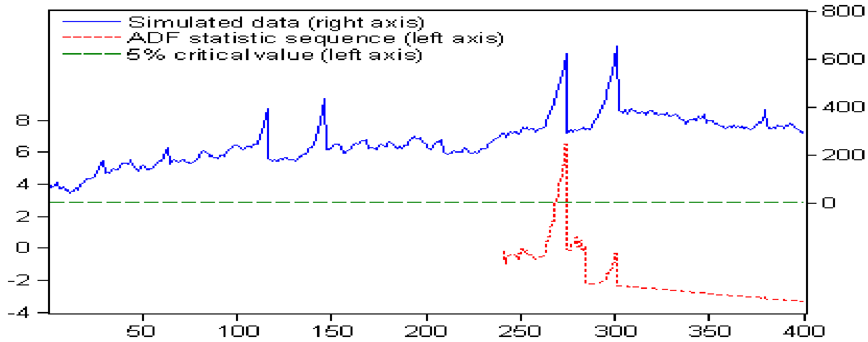
The smallest window size considered in the sup ADF test for the whole sample is 0.1, which contains 40 observations. The ADF statistic sequence of the sup ADF test is displayed in Figure 2 and the peak of the sequence is the defined sup ADF statistic. The sup ADF statistic of the simulated data series is 1.20, which is smaller than the 5% critical value of the sup ADF statistic 2.85. Therefore, we conclude that there are no bubbles in this sample.

Fig. 2. The sup ADF test

(a) Sample: 1 to 400 with $r_0 = 0.1$



(b) Sample: 201 to 400 with $r_0 = 0.2$



Suppose the sup ADF test starts from the 201^{th} observation, which is right before the two largest spires. The smallest regression window also contains 40 observations ($r_0 = 0.2$). The respective ADF statistic sequence is displayed in Figure 2. The sup ADF statistic obtained from this sample is 6.47 and it is greater than 2.85. [5] Thus, we confirm the existence of bubbles.

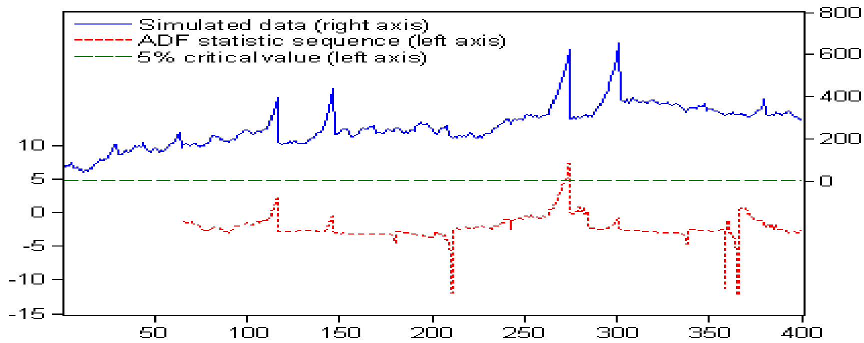
As we can see, the sup ADF test fails to reveal the existence of bubbles when the whole sample is utilized, whereas by re-selecting the starting point of the sample to exclude spires before the largest one, it manages to confirm the existence of bubbles. Both experiments above can be viewed as special cases of the generalized sup ADF test, where the sample starting points are fixed. In the first experiment,

[5] From 200 observations, the 5% critical value obtained from Monte Carlo simulation with 5,000 replications is 2.68 ($r_0 = 0.2$).

the sample starting point of the generalized sup ADF test r_1 is set to 0. The sample starting point r_1 of the second experiment is fixed at 0.502. The conflicting results we obtained from these two experiments also demonstrate the importance of a varying starting point in the generalized sup ADF test.

We then apply the generalized sup ADF test to the simulated asset prices. The optimum window size is 65 upon considering the sequence of window size from 40 to 120.[6] Figure 3 illustrates the ADF statistic sequence with the optimum window size. The generalized sup ADF statistic of the simulated data is 7.35, which is greater than the 5% critical value 4.75. It implies that the generalized sup ADF test find evidence of bubble existence. Compared to the sup ADF test, the generalized sup ADF identifies bubbles without mining the sample starting point, which is an obvious improvement.[7]

Fig. 3. The generalized sup ADF test



Notice that both the sup ADF test and the generalized sup ADF test are tests of the explosivity of the largest spire within the sample range. Therefore, only the peak of their ADF statistic sequences can be compared to the respective 5% critical value. To explore the significance of the second highest ADF statistics in the sequences,

[6] To improve the computation speed, the maximum window size is set 0.3 (=120/400) instead of 1. The maximum window size normally can be set according to the feature of respective data series.

[7] We observe similar phenomenon from the alternative data generating process where the logarithmic dividend is a random walk with drift. Parameters in the alternative data generating process are set as in Hall et al. (1999): $B_0 = 0.5, \alpha = 1, \pi = 0.85, \zeta = 0.5, \rho = 0.952, \tau = 0.05, \mu = 0.013, D_0 = 0.26$, $\sigma_D^2 = 0.016$, and $P_t = P_t^f + 250B_t$.

we need critical values for the second largest ADF statistics or we have to exclude the largest spire from the sample range. The same argument applies to other spires in the sample range. Therefore, we cannot conclude that the rest of the spires in the simulated data series are not explosive simply based on the sup ADF test or the generalized sup ADF test.

5 POWER AND SIZE COMPARISON

This section compares the powers and sizes of the sup ADF test and the generalized sup ADF test. The basic idea of power comparison is to repeatedly generate data series with explosive and periodically collapsing properties, and compare the proportion of simulations for which the method draws the right conclusion that there exists bubbles. For the size comparison, we need to repeatedly generate a data series without bubbles, and compare the proportions of the simulations for which the method draws the wrong conclusion that there exists bubbles.

The data generating process of the power comparison is the same as the simulation in section 4, which is the summation of a market fundamental component and a bubble component. The data generating process of the size comparison is based on equation (5) and equation (6), which is the market fundamental component. The critical values of the sup ADF test and the generalized sup ADF test are displayed in Table 2. The number of iterations for power and size calculations are 1,000. The explosive alternative is tested at the 5% significance level. The sample size equals 400. The smallest samples of both tests are set the same as these in calculating respective critical values, which are both set to have 40 observations.

Table 3. Powers and Sizes of the ADF tests (obs.=400)

	ADF	SADF	GSADF
Power	0.02	0.67	0.85
Size (5%)	0.022	0.018	0.023

Note: the number of iterations for power and size calculation equals 1000. The smallest sample has 40 observations ($r_0 = 0.1$).

Table 3 depicts the calculated powers and sizes of these methods. As shown in Evans (1991), the conventional ADF test performs poorly in the presence of periodically collapsing bubbles. We confirm the power improvement of the sup ADF test as in PWY. The generalized sup ADF test which is proposed to overcome the pitfall of the sup ADF test increases the testing power from 0.67 to 0.85. Furthermore, there is no significant difference in the sizes of these three different methods at the 5% level. Thus, we conclude that the generalized sup ADF test performs better in revealing the existence of explosive behavior. [8]

Table 4. Powers and Sizes of the ADF tests (obs.=200)

	ADF	SADF	GSADF
Power	0.03	0.50	0.62
Size (5%)	0.014	0.021	0.026

Note: The 5% critical values are obtained from 5,000 times Monte Carlo simulations with sample size 200, which are 1.09, 2.82 and 4.30 for the ADF test, sup ADF test and the generalized sup ADF test respectively. The number of iterations for power and size calculation equals 1000. The smallest sample has 40 observations ($r_0 = 0.2$).

Similar power and size patterns are observed in Table 4, where the sample size is 200. The powers of the sup ADF test and the generalized sup ADF test decrease with the sample size; nevertheless, the power of the generalized sup ADF test remains higher than the sup ADF test. The size difference between the sup ADF test and the generalized sup ADF test again is not significant.

6 APPLICATION: HONG KONG STOCK MARKET

To highlight the differences between the sup ADF test and the generalized sup ADF test, we investigate the presence of bubbles in Hong Kong stock market. Many papers have studied the evidence of bubble existence in Hong Kong stock market (e.g., Sornette and Johansen (2001), Zhou and Sornette (2003) and Cajueiro and Tabak

[8] Powers of the ADF test, the sup ADF test and the generalized sup ADF test with sample size 400 under the alternative data generating process, where we assume that the logarithmic dividend is a random walk, are 0.142, 0.914 and 0.957 respectively.

(2006), among others). Although the sample periods and the sample frequencies examined by these papers are different, most of them find evidences of of bubble existence.

Our data is sampled monthly over the period from October 1980 to April 2009, constituting 343 observations. The data comprises the Hang Seng Index (HSI) and the consumption price index. The Hang Seng Index is downloaded from Datastream International. The consumption price index (October 2004-September 2005 =100) is obtained from the Hong Kong Monetary Authority. The consumption price index is used to convert the stock prices into real series.

Fig. 4. Real Hang Seng Index (normalized to 100 at the beginning of data series)

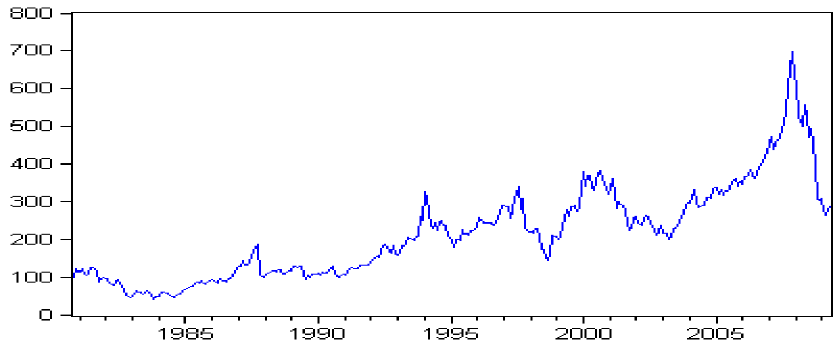


Figure 4 illustrates the behavior of the real Hang Seng Index (normalized to 100 at the beginning of the data series) during the data period. As we can see from the graph, the real Hang Seng Index fluctuates throughout the sample range. It is extremely volatile in the nine years spanning from 1994 to 2002 due to the 1997 Asian financial crisis, and a considerable increase that occurred over the period from April 2003 to November 2007. The peak of this increase was 6.97 times bigger than the starting point of the series. The real Hang Seng Index, then dropped quickly so that by March 2009 it was only 2.65 times that of the starting point. This last change is obviously related to the subprime crisis.

We apply the sup ADF test and the generalized sup ADF test to the logarithmic real HSI. Table 5 presents critical values for these two tests and these were obtained from 5,000 times Monte Carlo simulation under the null hypothesis that the true

Table 5. The sup ADF test and the generalized sup ADF test of the logarithmic real Hang Seng Index

	SADF	GSADF
Log real HSI	1.25	5.02
Stationary Alternative $H_1 : \beta < 0$		
1%	-1.24	1.19
5%	-0.80	1.65
10%	-0.57	1.94
Explosive Alternative $H_1 : \beta > 0$		
10%	1.55	4.57
5%	1.94	5.19
1%	2.85	6.71

Note: The optimum window of the logarithmic real Hang Seng Index has 37 observations. Critical values of both tests are obtained from 5,000 Monte Carlo simulations with sample size 343 under the null hypothesis that the true process is a random walk without drift. The smallest window is set to have 34 observations.

process is a random walk without drift. In both performing the ADF regressions and calculating the critical values, the smallest sample considered has 34 observations.

From Table 5, the sup ADF statistic of the logarithmic real Hang Seng Index is 1.25, which is smaller than the 10% right-tail critical value 1.55 and greater than the 10% critical value for the stationary alternative of the sup ADF test -0.57. Based on the sup ADF test, we conclude that the logarithmic real Hang Seng Index has a unit root. The generalized sup ADF statistic of the logarithmic real Hang Seng index is 5.02, which is greater than the respective 10% critical value of the explosive alternative, 4.57. This suggests that the Hang Seng Index is explosive based on the generalized sup ADF test, which contradicts the result from the sup ADF test.

7 CONCLUSION

The sup ADF test, also referred to as the forward recursive ADF test, implements the ADF test repeatedly on a sequence of forward expanding samples. The generalized sup ADF test can be viewed as a rolling window ADF test with a double-sup optimum window selection criteria [9]. That is, we select an optimum window size using the double-sup criteria and implements the ADF test repeatedly on a sequence of samples, which moves the optimum window frame gradually toward the end of the sample. By experimenting on simulated asset prices, we show the pitfall associated with the sup ADF test – inability to find bubbles when there are many spires in the sample range. In contrast to the sup ADF test, the generalized sup ADF test is able to surmount this problem and we show that it significantly improves the power to finding bubbles.

We apply both the sup ADF test and generalized sup ADF test to the Hang Seng Index from October 1980 to April 2009. This series contained many spires before the prices soared in 2006 and subsequently crashed in 2007 (the well-known subprime crisis). This is similar to the simulated scenario that highlighted the pitfall of the sup ADF test and the test results are consistent with our expectation. The generalized sup ADF suggests that there is explosive behavior in the Hang Seng Index, whereas the sup ADF test does not.

REFERENCES

[1] O. J. Blanchard. Speculative bubbles, crashes and rational expectations. *Economics Letters*, 3:387–389, Nov 1979.

[2] D. O. Cajueiro and B. M. Tabak. Testing for rational bubbles in banking indices. *Physica A*, 366:365–376, 2006.

[3] J. Y. Campbell and R. J. Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228, 1989.

[4] W. W. Charemza and D. F. Deadman. Speculative bubbles with stochastic

[9] First, we calculate the sup value of the ADF statistic over the feasible ranges of the window starting points for a fixed window size. Then, we calculate the sup value of the sup ADF statistic over the feasible range of window sizes.

explosive roots: The failure of unit root testing. *Journal of Empirical Finance*, 2:153–163, 1995.

[5] B. T. Diba and H. I. Grossman. Explosive rational bubbles in stock prices? *The American Economic Review*, 78(3):520–530, Jun 1988.

[6] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, June 1979.

[7] G. W. Evans. Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81(4):922–930, Sep. 1991.

[8] K. A. Froot and M. Obstfeld. Intrinsic bubbles: The case of stock prices. *American Economic Review.*, 81:1189–1214, 1991.

[9] R. S. Gurkaynak. Econometric tests of asset price bubbles: Taking stock. *Journal of Economic Surveys*, 22(1):166–186, 2008.

[10] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, March 1989.

[11] J. D. Hamilton. *Time Series Analysis.* Princeton University Press, 1 edition, 1994.

[12] S. F. LeRoy and R. D. Porter. The present-value relation: Tests based on implied variance bounds. *Econometrica*, 49:555–574, 1981.

[13] N. Li. Intangible capital and stock prices. *Singapore National University, Working Paper*, 2005.

[14] S. Ng and P. Perron. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554, 2001.

[15] P. C. B. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[16] P. C. B. Phillips, Y. Wu, and J. Yu. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review, forthcoming*, 2009.

[17] R. J. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review.*, 71:421–436, 1981.

[18] D. Sornette and A. Johansen. Significance of log-periodic precursors to financial crashes. *Quantitative Finance*, 1:452–471, 2001.

[19] J. Tirole. On the possibility of speculation under rational expectations. *Econometrica*, 50(5):1163–1181, Sep 1982.

[20] K. D. West. A specification test for speculative bubbles. *Quarterly Journal of Economics.*, 102:553–580, 1987.

[21] Y. Wu. Rational bubbles in the stock market: Accounting for the U.S. stock-price volatility. *Economic lnquiry*, XXXV:309–319, 1997.

[22] W. Zhou and D. Sornette. Evidence of worldwide stock market log-periodic anti-bubble since mid-2000. *Physica A*, 330:543 583, 2003.

[23] E. Zivot and D. W. K. Andrews. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3):251–270, Jul 1992.

A A SIMPLE ASSET PRICING MODEL

Consider a simple asset pricing model where risk-neutral investors choose between consumption and holding a risky asset. Suppose there exists a risk-free interest rate r_f, the period-to-period arbitrage condition for the asset is [10]

$$P_t = \rho \mathbb{E}_t \left(P_{t+1} + D_{t+1} \right) \tag{A.1}$$

where $\rho = \left(1 + r_f\right)^{-1}$, P_t is the after-dividend price of the asset (i.e stock price), and D_t is the payoff received from the asset (i.e. dividend).

Iterating equation (A.1) forward, we can obtain

$$P_t = \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t D_{t+j} + \lim_{j \to \infty} \rho^j \mathbb{E}_t \left(P_{t+j} \right). \tag{A.2}$$

The first component of equation (A.2) is defined as the market fundamental of the asset prices P_t^f and the second component is defined as the bubble component B_t. Those are,

$$P_t^f = \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t D_{t+j} \tag{A.3}$$

$$B_t = \lim_{j \to \infty} \rho^j \mathbb{E}_t P_{t+j} \tag{A.4}$$

The conditional expectation of ρB_{t+1} is

$$\begin{aligned} \mathbb{E}_t \left(\rho B_{t+1} \right) &= \lim_{j \to \infty} \rho^{j+1} \mathbb{E}_t \left[\mathbb{E}_{t+1} P_{t+1+j} \right] \\ &= \lim_{j \to \infty} \rho^{j+1} \mathbb{E}_t P_{t+1+j} \\ &= \lim_{k \to \infty} \rho^k \mathbb{E}_t P_{t+k} = B_t \end{aligned}$$

Therefore, we have $\mathbb{E}_t \left(B_{t+1} \right) = \left(1 + r_f\right) B_t$.

[10] By imposing these restrictions, we illustrate a simplified version of the cointegration relationship between stock prices and dividends. For a general description see Campbell and Shiller (1989).

A.1 THE COINTEGRATION RELATIONSHIP

Define $p_t = \ln P_t$ and $d_t = \ln D_t$, equation (A.1) can be rewritten as

$$e^{p_t} = \rho \mathbb{E}_t \left[e^{p_{t+1}} + e^{d_{t+1}} \right]$$

Applying the Taylor series expansion at the sample mean $\bar{p}$ and $\bar{d}$,

$$p_t \approx \kappa + \rho \mathbb{E}_t \left[p_{t+1} + e^{\bar{d}-\bar{p}} d_{t+1} \right]$$

where $\kappa = (\rho - 1)(1 - \bar{p}) + \rho e^{\bar{d}-\bar{p}} \left(1 - \bar{d}\right)$. By iterating forward, we can get:

$$p_t \approx \kappa \left(1 + \rho + \rho^2 + \ldots\right) + \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j}\right) + \lim_{j \to \infty} \rho^j \mathbb{E}_t \left(p_{t+j}\right)$$

Since $|\rho| < 1$, so we can get

$$p_t \approx \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j}\right) + \lim_{j \to \infty} \rho^j \mathbb{E}_t \left(p_{t+j}\right)$$

If there is no bubble in the market $b_t = 0$; $p_t = p_t^f$. Then,

$$p_t = p_t^f \approx \frac{\kappa}{1-\rho} + \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j}\right)$$

Multiplying both sides by $(1 - \rho)$,

$$\begin{aligned}
(1-\rho) p_t^f &= \kappa + (1-\rho) \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j}\right) \\
&= \kappa + \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j}\right) - \sum_{j=2}^{\infty} \rho^j \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+j-1}\right) \\
&= \kappa + \rho \mathbb{E}_t \left(e^{\bar{d}-\bar{p}} d_{t+1}\right) + \sum_{j=2}^{\infty} \rho^j \mathbb{E}_t \left[e^{\bar{d}-\bar{p}} \left(d_{t+j} - d_{t+j-1}\right)\right] \\
&= \kappa + \rho e^{\bar{d}-\bar{p}} d_t + \rho \mathbb{E}_t \left[e^{\bar{d}-\bar{p}} \triangle d_{t+1}\right] + \sum_{j=2}^{\infty} \rho^j \mathbb{E}_t \left[e^{\bar{d}-\bar{p}} \triangle d_{t+j}\right]
\end{aligned}$$

The last equation can be rewritten as

$$(1-\rho) p_t^f - \rho e^{\bar{d}-\bar{p}} d_t = \kappa + e^{\bar{d}-\bar{p}} \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left[\triangle d_{t+j}\right].$$

Suppose dividend D_t follows a unit root process with drift

$$D_t = \mu + D_{t-1} + \varepsilon_t, \varepsilon_t \sim N\left(0, \sigma_D^2\right).$$

Iterating backwards, the dividend process can be rewritten as

$$D_{t+j} = j\mu + D_t + \sum_{i=1}^{j} \varepsilon_{t+i}. \tag{A.5}$$

Combining equation (A.3) and equation (A.5), the market fundamental component of asset price can be written as

$$P_t^f = \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t D_{t+j} = \mu \sum_{j=1}^{\infty} j\rho^j + \sum_{j=1}^{\infty} \rho^j D_t = \frac{\mu\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho} D_t.$$

Suppose the logarithmic dividend follows a unit root process with drift

$$\ln D_t = \mu + \ln D_{t-1} + \varepsilon_t, \varepsilon_t \sim N\left(0, \sigma_d^2\right).$$

Iterating backwards, the dividend process can be rewritten as

$$D_{t+j} = \exp\left(j\mu + \textstyle\sum_{i=1}^{j} \varepsilon_{t+i}\right) D_t \tag{A.6}$$

Combining equation (A.3) and equation (A.6), the market fundamental component of asset price can be written as

$$\begin{aligned} P_t^f &= \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t D_{t+j} = \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t \left[\exp\left(j\mu + \textstyle\sum_{i=1}^{j} \varepsilon_{t+i}\right) D_t\right] \\ &= \sum_{j=1}^{\infty} \rho^j \exp\left(j\mu + \frac{1}{2} j\sigma_d^2\right) D_t \\ &= \frac{\rho \exp\left(\mu + \frac{1}{2}\sigma_d^2\right)}{1 - \rho \exp\left(\mu + \frac{1}{2}\sigma_d^2\right)} D_t \end{aligned}$$

B PROPOSITIONS AND PROOFS

Proposition 1 *Let* $u_t = \psi(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\psi\varepsilon_{t-j}$*, where* $\Sigma_{j=0}^{\infty}j.|\psi_j| < \infty$ *and* $\{\varepsilon_t\}$ *is an i.i.d sequence with mean zero, variance* σ^2 *and finite fourth moment. Define* $y_t = \sum_{s=1}^{t}u_s$ *with* $y_0 = 0$*,* $r_2 = r_1 + r_w$ *and* $\gamma_j \equiv E(u_t u_{t-j}) = \sigma^2\sum_{s=0}^{\infty}\psi_s\psi_{s+j}$ *for* $j = 0, 1, 2\cdots$ *. Then, we can calculate that*

$$(a)\ T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t \xrightarrow{L} \sigma^2\psi(1)\left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2}r_w\right]$$

$$(b)\ T^{-1/2}\sum_{t=[Tr_1]}^{[Tr_2]} \varepsilon_t \xrightarrow{L} \sigma\left[W(r_2) - W(r_1)\right]$$

$$(c)\ T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} \xrightarrow{L} \psi(1)\,\sigma\int_{r_1}^{r_2} W(r)\,dr$$

$$(d)\ T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}^2 \xrightarrow{L} \sigma^2\left[\psi(1)\right]^2\int_{r_1}^{r_2}\left[W(r)\right]^2 dr$$

$$(e)\ T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} u_{t-j} \xrightarrow{p} 0$$

$$(f)\ T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{p} 0, j = 0, 1, \cdots$$

where W *is the standard Wiener process.*

Claim 1 $y_t = \psi(1)\sum_{s=1}^{t}\varepsilon_s + \eta_t - \eta_0$*, where* $\eta_t = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$*,* $\eta_0 = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{-j}$ *and* $\alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots)$*, which is absolutely summable.*

Please refer to Hamilton (1994, ch. 17 pp 534-535) for the proof.

Claim 2 $\frac{1}{T}\sum_{t=[Tr_1]}^{[Tr_2]}\varepsilon_t^2 \xrightarrow{p} r_w.\sigma^2$.

Since $\frac{[Tr_w]}{T} \to r_w$ as $N \to \infty$, so by the law of large numbers,

$$\frac{1}{T}\sum_{t=[Tr_1]}^{[Tr_2]}\varepsilon_t^2 = \frac{[Tr_w]}{T}.\frac{1}{[Tr_w]}\sum_{t=[Tr_1]}^{[Tr_2]}\varepsilon_t^2 \xrightarrow{p} r_w\sigma^2.$$

Claim 3 $\sqrt{T}X_T(.) \xrightarrow{L} \sigma W(.)$*, where* $X_T(r)$ *is the sample mean of the first* r^{th} *fraction of observations* $\{\varepsilon_t\}_{t=1}^{T}$*.*

Define a random walk process $Z_{t-1} = \sum_{s=1}^{t-1} \varepsilon_s$, then $X_T(r) = \frac{1}{T}\sum_{s=1}^{[Tr]} \varepsilon_s = \frac{1}{T} Z_{[Tr]}$. For any given realization, $X_T(r)$ is a step function in r, with

$$X_T(r) = \begin{cases} 0 \quad \text{for } 0 \le r < \frac{1}{T} \\ \frac{Z_1}{T} \ \text{ for } \frac{1}{T} \le r < \frac{2}{T} \\ \cdots \\ \frac{Z_T}{T} \qquad \text{for } r = 1 \end{cases}.$$

By the definition of $X_T(r)$,

$$\sqrt{T} X_T(r) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[Tr]} \varepsilon_s = \frac{\sqrt{[Tr]}}{\sqrt{T}} \frac{1}{\sqrt{[Tr]}} \sum_{s=1}^{[Tr]} \varepsilon_s.$$

By the central limit theorem, $\frac{1}{\sqrt{[Tr]}} \sum_{s=1}^{[Tr]} \varepsilon_s \xrightarrow{L} N(0, \sigma^2)$. Since $\frac{\sqrt{[Tr]}}{\sqrt{T}} \to \sqrt{r}$ when $T \to \infty$, so $\sqrt{T} X_T(r) \xrightarrow{L} N(0, r\sigma^2)$. By the functional central limit theorem, we have

$$\sqrt{T}\frac{X_T(.)}{\sigma} \xrightarrow{L} W(.) \ \text{ or } \sqrt{T} X_T(.) \xrightarrow{L} \sigma W(.)$$

where W is the standard Wiener process.

Claim 4 *$\frac{1}{T} Z_{[Tr_1]}^2 \xrightarrow{L} \sigma^2 [W(r_1)]^2$ and $\frac{1}{T} Z_{[Tr_2]}^2 \xrightarrow{L} \sigma^2 [W(r_2)]^2$, where $Z_{[Tr_1]} = \sum_{s=1}^{[Tr_1]} \varepsilon_s$ and $Z_{[Tr_2]} = \sum_{s=1}^{[Tr_2]} \varepsilon_s$.*

Define $S_T(r) = \left[\sqrt{T} X_T(r)\right]^2$, which can be written as

$$S_T(r) = \begin{cases} 0 \quad \text{for } 0 \le r < \frac{1}{T} \\ \frac{Z_1^2}{T} \ \text{ for } \frac{1}{T} \le r < \frac{2}{T} \\ \cdots \\ \frac{Z_T^2}{T} \qquad \text{for } r = 1 \end{cases} \tag{B.1}$$

By Claim 3 and the continuous mapping theorem, we have

$$\frac{1}{T}Z^2_{[Tr_1]} = S_T(r_1) \xrightarrow{L} \sigma^2 [W(r_1)]^2$$
$$\frac{1}{T}Z^2_{[Tr_2]} = S_T(r_2) \xrightarrow{L} \sigma^2 [W(r_2)]^2 .$$

Claim 5 $T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]}\sum_{s=1}^{t-1}\varepsilon_s.\varepsilon_t \xrightarrow{L} \sigma^2\left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2}r_w\right].$

The definition of Z_t implies that $Z_t = Z_{t-1} + \varepsilon_t$. Summing $Z_{t-1}\varepsilon_t$ from $[Tr_1]$ to $[Tr_2]$ and dividing by T,

$$\frac{1}{T}\sum_{t=[Tr_1]}^{[Tr_2]} Z_{t-1}\varepsilon_t = \frac{1}{2}\frac{1}{T}Z^2_{[Tr_2]} - \frac{1}{2}\frac{1}{T}Z^2_{[Tr_1]} - \frac{1}{2}\frac{1}{T}\sum_{t=[Tr_1]}^{[Tr_2]}\varepsilon_t^2.$$

Therefore, combining Claim 4 and Claim 2, we can get

$$\frac{1}{T}\sum_{t=[Tr_1]}^{[Tr_2]} Z_{t-1}\varepsilon_t \xrightarrow{L} \frac{1}{2}\sigma^2\left\{[W(r_2)]^2 - [W(r_1)]^2 - r_w\right\} = \sigma^2\left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2}r_w\right].$$

Claim 6 $T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]}(\eta_{t-1} - \eta_0)\varepsilon_t \xrightarrow{p} 0$ *as* $T \to \infty$.

Assumptions in the proposition ensure that $\{(\eta_{t-1} - \eta_0)\varepsilon_t\}_{t=1}^{\infty}$ is a martingale difference sequence with finite variance, so we complete the proof.

Claim 7 $1/\sqrt{T}\left(\eta_{[Tr]} - \eta_0\right) \to 0$ *as* $T \to \infty$.

Assumptions in the proposition ensure that $\{\eta_{t-1} - \eta_0\}_{t=1}^{\infty}$ is a martingale difference sequences with finite variance, so we complete the proof.

(a) From Claim 1, we have

$$\begin{aligned} T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t &= T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]}\left(\psi(1)\sum_{s=1}^{t-1}\varepsilon_s + \eta_{t-1} - \eta_0\right)\varepsilon_t \\ &= \psi(1)\,T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t + T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]}(\eta_{t-1} - \eta_0)\,\varepsilon_t. \end{aligned}$$

Therefore, by Claim 5 and Claim 6, we have

$$T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t \xrightarrow{L} \sigma^2\psi(1)\left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2}r_w\right].$$

(b) From Claim 3,

$$T^{-1/2}\sum_{t=[Tr_1]}^{[Tr_2]} \varepsilon_t = \sqrt{T}\left[X_T(r_2) - X_T(r_1)\right] \xrightarrow{L} \sigma\left[W(r_2) - W(r_1)\right].$$

(c) Define $M_T(r) = 1/T\sum_{s=1}^{[T\tilde{r}]} u_s$. From the definition $y_t = \sum_{s=1}^{t} u_s$, we can write $M_T(r)$ as

$$M_T(r) = \begin{cases} 0 \quad \text{for } 0 \le r < \frac{1}{T} \\ \frac{y_1}{T} \quad \text{for } \frac{1}{T} \le r < \frac{2}{T} \\ \cdots \\ \frac{y_T}{T} \qquad \text{for } r = 1 \end{cases}.$$

It follows that

$$\begin{aligned} \sqrt{T}M_T(r) &= 1/\sqrt{T}y_{[Tr]} \\ &= 1/\sqrt{T}\left[\psi(1)\sum_{s=1}^{[Tr]}\varepsilon_s + \eta_{[Tr]} - \eta_0\right] \text{ (from Claim 1)} \\ &= \psi(1)\left(1/\sqrt{T}\right)\sum_{s=1}^{[Tr]}\varepsilon_s + 1/\sqrt{T}\left(\eta_{[Tr]} - \eta_0\right) \end{aligned}$$

From Claim 3 and Claim 7, we can get

$$\sqrt{T}M_T(r) \xrightarrow{L} \psi(1)\sigma W(r).$$

Integrating $\sqrt{T}M_T(r)$ from r_1 to r_2,

$$\int_{r_1}^{r_2}\sqrt{T}M_T(r)\,dr = \sqrt{T}.\frac{1}{T}\left(\frac{y_{[Tr_1]}}{T} + \cdots + \frac{y_{[Tr_2]}}{T}\right)$$
$$= T^{-3/2}\sum_{s=[Tr_1]}^{[Tr_2]} y_s \xrightarrow{L} \psi(1)\sigma\int_{r_1}^{r_2} W(r)\,dr.$$

(d) Define $N_T(r) = \left[\sqrt{T} M_T(r)\right]^2$. We can write $N_T(r)$ as

$$N_T(r) = \begin{cases} 0 \quad \text{for } 0 \leq r < \frac{1}{T} \\ \frac{y_1^2}{T} \ \text{ for } \frac{1}{T} \leq r < \frac{2}{T} \\ \cdots \\ \frac{y_T^2}{T} \qquad \text{for } r = 1 \end{cases}$$

Integrating r from r_1 to r_2,

$$\int_{r_1}^{r_2} N_T(r)\, dr = \frac{1}{T}\left(\frac{y_{[Tr_1]}^2}{T} + \cdots + \frac{y_{[Tr_2]}^2}{T}\right) = T^{-2} \sum_{s=[Tr_1]}^{[Tr_2]} y_s^2.$$

By the continuous mapping theorem,

$$T^{-2} \sum\nolimits_{t=[Tr_1]}^{[Tr_2]} y_t^2 \xrightarrow{L} \sigma^2 [\psi(1)]^2 \int_{r_1}^{r_2} [W(r)]^2\, dr.$$

(e) Since $\frac{[Tr_w]}{T} \to r_w$ as $N \to \infty$, so by the law of large numbers,$\sum$

$$T^{-1} \sum_{t=[Tr_1]}^{[Tr_2]} u_{t-j} = \frac{[Tr_w]}{T} [Tr_w]^{-1} \sum_{t=[Tr_1]}^{[Tr_2]} u_{t-j} \xrightarrow{p} r_w.\mathbb{E}(u_t) = 0.$$

Claim 8 $T^{-1} \sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} u_t \xrightarrow{L} \frac{1}{2}\psi(1)^2 \sigma^2 [W(r_2)]^2 - \frac{1}{2}(r_2 - r_1)\gamma_0$.

By definition of $y_t = \sum_{s=1}^t u_s$, we have

$$T^{-1} \sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} u_t = \frac{1}{2} T^{-1} y_{[Tr_2]}^2 - \frac{1}{2} T^{-1} \left(\sum_{s=1}^{[Tr_2]} u_s^2 - \sum_{s=1}^{[Tr_1]} u_s^2\right).$$

From part (d), $T^{-1} y_{[Tr_2]}^2 = \left[T^{-1/2}\left(u_1 + \cdots + u_{[Tr_2]}\right)\right]^2 \xrightarrow{L} \sigma^2 [W(r_2)]^2$. Further,more,

$$T^{-1} \sum_{s=1}^{[Tr_2]} u_s^2 = \frac{[Tr_2]}{T} \frac{1}{[Tr_2]} \sum_{s=1}^{[Tr_2]} u_s^2 \xrightarrow{p} r_2.\gamma_0$$

$$T^{-1} \sum_{s=1}^{[Tr_1]} u_s^2 = \frac{[Tr_1]}{T} \frac{1}{[Tr_1]} \sum_{s=1}^{[Tr_1]} u_s^2 \xrightarrow{p} r_1.\gamma_0.$$

Therefore,

$$T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_t \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r_2)\right]^2 - \frac{1}{2}(r_2-r_1)\gamma_0.$$

Claim 9 *For* $j = 1, 2, \cdots,$

$$T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r_2)\right]^2 - \frac{1}{2}(r_2-r_1)\gamma_0 + r_w\sum_{s=1}^{j}\gamma_{s-1}.$$

We observe that $y_{t-1} = y_{t-j-1} + \sum_{s=1}^{j} u_{t-s}$, which implies that

$$\frac{1}{T}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-1}u_{t-j} = \frac{1}{T}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-j-1}u_{t-j} + \frac{1}{T}\sum_{t=[Tr_1]+j}^{[Tr_2]}\sum_{s=1}^{j} u_{t-s}u_{t-j} \tag{B.2}$$

From Claim 8, we have

$$T^{-1}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-j-1}u_{t-j} \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r_2)\right]^2 - \frac{1}{2}(r_2-r_1)\gamma_0.$$

Since $T^{-1}\sum_{t=[Tr_1]+j}^{[Tr_2]}\sum_{s=1}^{j} u_{t-s}u_{t-j} \to r_w\sum_{s=1}^{j}\gamma_{s-1}$, equation B.2 converges to

$$T^{-1}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r_2)\right]^2 - \frac{1}{2}(r_2-r_1)\gamma_0 + r_w\sum_{s=1}^{j}\gamma_{s-1}.$$

Combining with the fact that $T^{-1}\sum_{t=[Tr_1]}^{[Tr_1]+j} y_{t-1}u_{t-j} \xrightarrow{p} 0$,

$$T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} \frac{1}{2}\sigma^2\left[W(r_2)\right]^2 - \frac{1}{2}(r_2-r_1)\gamma_0 + r_w\sum_{s=1}^{j}\gamma_{s-1}.$$

(f) From Claim 8 and Claim 9, we have

$$T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} = T^{-1/2}.T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{p} 0, \text{ for } j = 0, 1, \cdots.$$

Proposition 2 *Let* $u_t = \psi(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$*, where* $\Sigma_{j=0}^{\infty}j.|\psi_j| < \infty$ *and* $\{\varepsilon_t\}$ *is an i.i.d sequence with mean zero, variance* σ^2 *and finite fourth moment. Define* $y_t = \tilde{\alpha}t + \sum_{s=1}^{t} u_s$ *with* $y_0 = 0$*,* $r_2 = r_1 + r_w$ *and* $\gamma_j \equiv E(u_t t_{t-j}) = \sigma^2 \sum_{s=0}^{\infty}\psi_s\psi_{s+j}$ *for* $j = 0, 1, 2\cdots$*. Then, we can calculate that*

$$(a)\ T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t \xrightarrow{p} T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]}\tilde{\alpha}(t-1)\varepsilon_t$$

$$(b)\ T^{-1/2}\sum_{t=[Tr_1]}^{[Tr_2]}\varepsilon_t \xrightarrow{L} \sigma[W(r_2) - W(r_1)]$$

$$(c)\ T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} \xrightarrow{p} \frac{\tilde{\alpha}}{2}r_w(r_1 + r_2)$$

$$(d)\ T^{-3}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}^2 \xrightarrow{p} \frac{\tilde{\alpha}^2}{3}\left(r_2^3 - r_1^3\right)$$

$$(e)\ T^{-1}\sum_{t=[Tr_1]}^{[Tr_2]} u_{t-j} \xrightarrow{p} 0$$

$$(f)\ T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{p} 0, , j = 0, 1, \cdots$$

where W is the standard Wiener process.

Claim 10 $\xi_t = \Sigma_{s=1}^{t}u_s = \psi(1)\sum_{s=1}^{t}\varepsilon_s + \eta_t - \eta_0$*, where* $\eta_t = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{t-j}$*,* $\eta_0 = \sum_{j=0}^{\infty}\alpha_j\varepsilon_{-j}$ *and* $\alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots)$*, which is absolutely summable.*

The proof is the same as Claim 1.

(a) From Claim 10, we have

$$\begin{aligned} T^{-3/2}\sum\nolimits_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t &= T^{-3/2}\sum\nolimits_{t=[Tr_1]}^{[Tr_2]}\left(\tilde{\alpha}(t-1) + \psi(1)\sum_{s=1}^{t-1}\varepsilon_s + \eta_{t-1} - \eta_0\right)\varepsilon_t \\ &= T^{-3/2}\sum\nolimits_{t=[Tr_1]}^{[Tr_2]}\tilde{\alpha}(t-1)\varepsilon_t + \psi(1)T^{-3/2}\sum\nolimits_{t=[Tr_1]}^{[Tr_2]}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t \\ &+ T^{-3/2}\sum\nolimits_{t=[Tr_1]}^{[Tr_2]}(\eta_{t-1} - \eta_0)\varepsilon_t. \end{aligned}$$

Therefore, by Claim 5 and Claim 6, we have

$$T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}\varepsilon_t \xrightarrow{p} T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]}\tilde{\alpha}(t-1)\varepsilon_t.$$

(b) The proof is the same as (b) in Proposition 1.

(c) Define $M_T(r) = 1/T \sum_{s=1}^{[Tr]} u_s$. Let $\xi_t = \sum_{s=1}^{t} u_s$, we can write $M_T(r)$ as

$$M_T(r) = \begin{cases} 0 \quad \text{for } 0 \leq r < \frac{1}{T} \\ \frac{\xi_1}{T} \quad \text{for } \frac{1}{T} \leq r < \frac{2}{T} \\ \cdots \\ \frac{\xi_T}{T} \qquad \text{for } r = 1 \end{cases}.$$

It follows that

$$\begin{aligned} \sqrt{T} M_T(r) &= 1/\sqrt{T} \xi_{[Tr]} \\ &= 1/\sqrt{T} \left[\psi(1) \sum_{s=1}^{[Tr]} \varepsilon_s + \eta_{[Tr]} - \eta_0 \right] \text{ (from Claim 10)} \\ &= \psi(1) \left(1/\sqrt{T} \right) \sum_{s=1}^{[Tr]} \varepsilon_s + 1/\sqrt{T} \left(\eta_{[Tr]} - \eta_0 \right) \end{aligned}$$

From Claim 3 and Claim 7, we can get

$$\sqrt{T} M_T(r) \xrightarrow{L} \psi(1) \sigma W(r).$$

Integrating $\sqrt{T} M_T(r)$ from r_1 to r_2,

$$\begin{aligned} \int_{r_1}^{r_2} \sqrt{T} M_T(r) \, dr &= \sqrt{T}. \frac{1}{T} \left(\frac{\xi_{[Tr_1]}}{T} + \cdots + \frac{\xi_{[Tr_2]}}{T} \right) \\ = T^{-3/2} \sum_{t=[Tr_1]}^{[Tr_2]} \xi_t &\xrightarrow{L} \psi(1) \sigma \int_{r_1}^{r_2} W(r) \, dr. \end{aligned}$$

Since $y_t = \tilde{\alpha} t + \xi_t$,

$$\begin{aligned} T^{-2} \sum_{s=[Tr_1]}^{[Tr_2]} y_s &= T^{-2} \sum_{s=[Tr_1]}^{[Tr_2]} \tilde{\alpha} t + T^{-2} \sum_{t=[Tr_1]}^{[Tr_2]} \xi_t \\ &= \frac{\tilde{\alpha}}{2} \frac{[Tr_w]([Tr_2] + [Tr_1])}{T^2} t + T^{-2} \sum_{s=[Tr_1]}^{[Tr_2]} \xi_t \\ &\xrightarrow{p} \frac{\tilde{\alpha}}{2} r_w (r_1 + r_2) \end{aligned}$$

(d) Define $N_T(r) = \left[\sqrt{T} M_T(r)\right]^2$. We can write $N_T(r)$ as

$$N_T(r) = \begin{cases} 0 \quad \text{for } 0 \leq r < \frac{1}{T} \\ \frac{\xi_1^2}{T} \quad \text{for } \frac{1}{T} \leq r < \frac{2}{T} \\ \cdots \\ \frac{\xi_T^2}{T} \qquad \text{for } r = 1 \end{cases}$$

Integrating r from r_1 to r_2,

$$\int_{r_1}^{r_2} N_T(r)\, dr = \frac{1}{T}\left(\frac{\xi_{[Tr_1]}^2}{T} + \cdots + \frac{\xi_{[Tr_2]}^2}{T}\right) = T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} \xi_t^2.$$

By the continuous mapping theorem,

$$T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} \xi_t^2 \xrightarrow{L} \sigma^2 [\psi(1)]^2 \int_{r_1}^{r_2} [W(r)]^2\, dr.$$

Then,

$$T^{-3}\sum_{t=[Tr_1]}^{[Tr_2]} y_t^2 = \tilde{\alpha}^2 T^{-3}\sum_{t=[Tr_1]}^{[Tr_2]} t^2 + T^{-3}\sum_{t=[Tr_1]}^{[Tr_2]} \xi_t^2 + 2\tilde{\alpha}T^{-3}\sum_{t=[Tr_1]}^{[Tr_2]} t\xi_t \xrightarrow{p} \frac{\tilde{\alpha}^2}{3}\left(r_2^3 - r_1^3\right)$$

(e) The proof is the same as (e) in Proposition 1.

Claim 11 $T^{-3/2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_t \xrightarrow{L} 0.$

By definition of $y_t = \tilde{\alpha} + y_{t-1} + u_t$, we have

$$\begin{aligned}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_t &= \frac{1}{2}\sum_{t=[Tr_1]}^{[Tr_2]}\left[y_t^2 - y_{t-1}^2 - \tilde{\alpha}^2 - 2\tilde{\alpha}y_{t-1} - 2u_t\tilde{\alpha} - u_t^2\right] \\ &= \frac{1}{2}y_{[Tr_2]}^2 - \frac{1}{2}y_{[Tr_1]-1}^2 - \frac{1}{2}\sum_{t=[Tr_1]}^{[Tr_2]}\tilde{\alpha}^2 - \tilde{\alpha}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} \\ &\quad - \tilde{\alpha}\left(\sum_{s=1}^{[Tr_2]} u_s - \sum_{s=1}^{[Tr_1]} u_s\right) - \frac{1}{2}\left(\sum_{s=1}^{[Tr_2]} u_s^2 - \sum_{s=1}^{[Tr_1]} u_s^2\right).\end{aligned}$$

From part (d), we know $T^{-1}\xi^2_{[Tr_2]} = \left[T^{-1/2}\left(u_1 + \cdots + u_{[Tr_2]}\right)\right] \xrightarrow{L} \sigma^2 \left[W\left(r_2\right)\right]^2$. So,

$$T^{-2}y^2_{[Tr_2]} = T^{-2}\left\{\tilde{\alpha}^2\left[Tr_2\right]^2 + 2\tilde{\alpha}\left[Tr_2\right] + \xi^2_{[Tr_2]}\right\} \xrightarrow{p} \tilde{\alpha}^2 r_2^2.$$

Similarly, we can get

$$T^{-2}y^2_{[Tr_1]-1} \xrightarrow{p} \tilde{\alpha}^2 r_1^2.$$

From part (c), $T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1} \xrightarrow{p} \frac{\tilde{\alpha}}{2} r_w\left(r_1 + r_2\right)$. Also,

$$T^{-1}\sum_{s=1}^{[Tr_2]} u_s^2 = \frac{[Tr_2]}{T}\frac{1}{[Tr_2]}\sum_{s=1}^{[Tr_2]} u_s^2 \xrightarrow{p} r_2.\gamma_0$$
$$T^{-1}\sum_{s=1}^{[Tr_1]} u_s^2 = \frac{[Tr_1]}{T}\frac{1}{[Tr_1]}\sum_{s=1}^{[Tr_1]} u_s^2 \xrightarrow{p} r_1.\gamma_0$$

Therefore,

$$T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_t \xrightarrow{p} \tilde{\alpha}^2\left(r_1 + r_2\right)\left(r_2 - r_1 - r_w\right) = 0$$

Claim 12 *For* $j = 1, 2, \cdots$, $T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} 0.$

We observe that $y_{t-1} = \tilde{\alpha}\left(j+1\right) + y_{t-j-1} + \sum_{s=1}^{j} u_{t-s}$, which implies that

$$\begin{aligned} T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-1}u_{t-j} &= T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]}\left[\tilde{\alpha}\left(j+1\right) + y_{t-j-1} + \sum_{s=1}^{j} u_{t-s}\right] u_{t-j} \\ &= T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]} \tilde{\alpha}\left(j+1\right) u_{t-j} + T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-j-1}u_{t-j} \\ &\quad + T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]}\sum_{s=1}^{j} u_{t-s}u_{t-j} \end{aligned} \tag{B.3}$$

From Claim 11, we have

$$T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-j-1}u_{t-j} \xrightarrow{L} 0.$$

Since $T^{-1}\sum_{t=[Tr_1]+j}^{[Tr_2]}\sum_{s=1}^{j} u_{t-s}u_{t-j} \to r_w\sum_{s=1}^{j}\gamma_{s-1}$, equation (B.3) converges to

$$T^{-2}\sum_{t=[Tr_1]+j}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} 0$$

Combining with the fact that $T^{-2}\sum_{t=[Tr_1]}^{[Tr_1]+j} y_{t-1}u_{t-j} \xrightarrow{p} 0,$

$$T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} 0.$$

(f) From Claim 11, and Claim 12, for $j = 0, 1, \cdots$, we have

$$T^{-2}\sum_{t=[Tr_1]}^{[Tr_2]} y_{t-1}u_{t-j} \xrightarrow{L} 0.$$

C THE ASYMPTOTIC DISTRIBUTION OF THE GENERALIZED SUP ADF STATISTIC

Consider the ADF model

$$\Delta y_t = \sum_{k=1}^{p-1} \phi^k \Delta y_{t-k} + \alpha + \beta y_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim N(0, \sigma^2)$. Suppose the sample starts from the r_1 fraction of the total sample and ends at fraction r_2, where $r_1 \in [0, 1 - r_w]$, $r_2 = r_1 + r_w$, $r_w \in [r_0, 1]$ is the window size fraction and r_0 is the smallest fraction considered $(0 < r_0 < 1)$. The deviation of the OLS estimate $\hat{\theta}$ from the true value θ is given by

$$\hat{\theta} - \theta = \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'\right]^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t\right] \tag{C.1}$$

where $X_t = [u_{t-1}\ u_{t-2}\ \ldots\ u_{t-k}\ 1\ y_{t-1}]'$, $\theta = [\psi_1\ \psi_2\ \ldots \psi_k\ \alpha\ \beta]'$ and $[.]$ signifies the integer part of its argument.

C.1 TRUE PROCESS IS A RANDOM WALK WITHOUT DRIFT

Assume the intimal value $y_0 = 0$. Under the null hypothesis that $\alpha = \beta = 0$, we have $y_t = \sum_{s=1}^{t} u_s$, where $u_t = \left(1 - \phi^1 L - \phi^2 L^2 - \cdots - \phi^{p-1} L^{p-1}\right)^{-1} \varepsilon_t = \psi(L)\varepsilon_t$.

From (e) and (f) of Proposition 1, we know that the probability limit of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'$ is a block diagonal. Therefore, we only need to obtain the last 2×2 components of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'$ and the last 2×1 component of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t$ to calculate the ADF statistics, which are

$$\begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{bmatrix}$$

respectively, where Σ denoting summation over $t = [Tr_1], [Tr_1]+1, \cdots, [Tr_2]$. Based on proposition 2, the scaling matrix should be $\Upsilon_T = diag\left(\sqrt{T}, T\right)$. Pre-multiplying equation (C.1) by Υ_T, results in

$$\Upsilon_T \begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} - \beta \end{bmatrix} = \left\{ \Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t' \right]_{(-2)\times(-2)} \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t \right]_{(-2)\times 1} \right\}$$

Consider the matrix $\Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t' \right]_{(-2)\times(-2)} \Upsilon_T^{-1}$,

$$\begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} = \begin{bmatrix} T^{-1}\Sigma 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} r_w & \psi(1)\,\sigma \int_{r_1}^{r_2} W(r)\,dr \\ \psi(1)\,\sigma \int_{r_1}^{r_2} W(r)\,dr & \sigma^2 [\psi(1)]^2 \int_{r_1}^{r_2} [W(r)]^2\,dr \end{bmatrix}.$$

and the matrix $\Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t \right]_{(-2)\times 1}$,

$$\begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1}\varepsilon_t \end{bmatrix} = \begin{bmatrix} T^{-1/2}\Sigma \varepsilon_t \\ T^{-1}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma [W(r_2) - W(r_1)] \\ \sigma^2 \psi(1) \left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2} r_w \right] \end{bmatrix}$$

Under the null hypothesis that $\beta = 0$,

$$\begin{bmatrix} \sqrt{T}\hat{\alpha} \\ T\hat{\beta} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} r_w & A \\ A & B \end{bmatrix}^{-1} \begin{bmatrix} C \\ D \end{bmatrix} = \frac{1}{A^2 - r_w B} \begin{bmatrix} -B & A \\ A & -r_w \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix}$$

where

$$A = \psi(1)\,\sigma \int_{r_1}^{r_2} W(r)\,dr$$
$$B = \sigma^2 [\psi(1)]^2 \int_{r_1}^{r_2} [W(r)]^2\,dr$$
$$C = \sigma [W(r_2) - W(r_1)]$$
$$D = \sigma^2 \psi(1) \left[\int_{r_1}^{r_2} W(r)\,dW(r) - \frac{1}{2} r_w\right].$$

Therefore, the generalization of the Dickey-Fuller test when lagged changes of y are included in the regression is

$$T\hat{\beta}.\psi(1) \xrightarrow{L} \frac{r_w \left[\int_{r_1}^{r_2} W dW - \frac{1}{2} r_w\right] - \int_{r_1}^{r_2} W dr.\,[W(r_2) - W(r_1)]}{r_w \int_{r_1}^{r_2} W^2 dr - \left[\int_{r_1}^{r_2} W dr\right]^2}$$

To calculate the t-statistic of $\hat{\beta}$, we need to find the standard error of $\hat{\beta}$. Since the variance of $\hat{\theta}$ is

$$var\left(\hat{\theta}\right) = var\left\{\left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'\right]^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t\right]\right\}$$
$$= \sigma^2 \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'\right]^{-1}$$

Therefore, variance of $\hat{\beta}$ is σ^2 multiply the last element of $\left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'\right]^{-1}$. We know that

$$var\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \sigma^2 \begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix}^{-1}$$

so, the variances of $T\hat{\beta}$ can be calculated as follows:

$$var\left(\begin{bmatrix} \sqrt{T}\hat{\alpha} \\ T\hat{\beta} \end{bmatrix}\right) = var\left(\Upsilon_T \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right)$$

$$= \sigma^2 \left\{ \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \right\}^{-1}$$

$$= \sigma^2 \begin{bmatrix} T^{-1}\Sigma 1 & T^{-3/2}\Sigma y_{t-1} \\ T^{-3/2}\Sigma y_{t-1} & T^{-2}\Sigma y_{t-1}^2 \end{bmatrix}^{-1} \xrightarrow{L} \sigma^2 \begin{bmatrix} r_w & A \\ A & B \end{bmatrix}^{-1}$$

Hence, the t-statistic of $\hat{\beta}$ is

$$\frac{\hat{\beta}}{se\left(\hat{\beta}\right)} = \frac{T\hat{\beta}}{se\left(T\hat{\beta}\right)} \xrightarrow{L} \frac{AC - r_w D}{A^2 - r_w B} . \left(\frac{r_w \sigma^2}{r_w B - A^2} \right)^{-1/2}$$

$$= \frac{r_w \left[\int_{r_1}^{r_2} W dW - \frac{1}{2} r_w \right] - \int_{r_1}^{r_2} W dr . \left[W\left(r_2\right) - W\left(r_1\right) \right]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W^2 dr - \left[\int_{r_1}^{r_2} W dr \right]^2 \right\}^{1/2}}$$

The asymptotic distribution of the t-statistic of the generalized sup ADF statistic is

$$\sup_{r_w \in [r_0, 1]} \sup_{\substack{r_1 \in [0, 1 - r_w] \\ r_2 = r_1 + r_w}} \left\{ \frac{r_w \left[\int_{r_1}^{r_2} W dW - \frac{1}{2} r_w \right] - \int_{r_1}^{r_2} W dr . \left[W\left(r_2\right) - W\left(r_1\right) \right]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W^2 dr - \left[\int_{r_1}^{r_2} W dr \right]^2 \right\}^{1/2}} \right\}.$$

In the sup ADF test, the starting point r_1 is fixed at 0, so $r_2 = r_1 + r_w = r_w$. Therefore, the asymptotic distribution of the sup ADF statistic is

$$\sup_{r_w \in [r_0, 1]} \left\{ \frac{r_w \left[\int_0^{r_w} W dW - \frac{1}{2} r_w \right] - \int_0^{r_w} W dr . W\left(r_w\right)}{r_w^{1/2} \left\{ r_w \int_0^{r_w} \left[W\left(r\right)\right]^2 dr - \left[\int_0^{r_w} W\left(r\right) dr \right]^2 \right\}^{1/2}} \right\}.$$

C.2 TRUE PROCESS IS A RANDOM WALK WITH DRIFT

Assume the initial value $y_0 = 0$. Under the null hypothesis that $\beta = 0$, we have $\Delta y_t = \tilde{\alpha} + u_t = \tilde{\alpha} t + \sum_{s=1}^{t} u_s$, where $u_t = \left(1 - \phi^1 L - \phi^2 L^2 - \cdots - \phi^{p-1} L^{p-1}\right)^{-1} \varepsilon_t = \psi\left(L\right) \varepsilon_t$ and $\tilde{\alpha} = \alpha \left(1 - \phi - \phi^2 - \cdots - \phi^{p-1}\right)^{-1} = \psi\left(1\right) \alpha$.

From (e) and (f) of Proposition 1, we know that the probability limit of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'$

is a block diagonal. Therefore, we only need to obtain the last 2×2 components of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t'$ and the last 2×1 component of $\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t$ to calculate the ADF statistics, which are

$$\begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1} \varepsilon_t \end{bmatrix}$$

respectively, where Σ denoting summation over $t = [Tr_1], [Tr_1]+1, \cdots, [Tr_2]$. Based on proposition 1, the scaling matrix should be $\Upsilon_T = diag\left(T^{1/2}, T^{\frac{3}{2}}\right)$. Pre-multiplying equation (C.1) by Υ_T, results in

$$\Upsilon_T \begin{bmatrix} \widehat{\alpha} - \alpha \\ \widehat{\beta} - \beta \end{bmatrix} = \left\{ \Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t' \right]_{(-2)\times(-2)} \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t \right]_{(-2)\times 1} \right\}$$

Consider the matrix $\Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t X_t' \right]_{(-2)\times(-2)} \Upsilon_T^{-1}$,

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma 1 & \Sigma y_{t-1} \\ \Sigma y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix} = \begin{bmatrix} T^{-1}\Sigma 1 & T^{-2}\Sigma y_{t-1} \\ T^{-2}\Sigma y_{t-1} & T^{-3}\Sigma y_{t-1}^2 \end{bmatrix}$$

$$\xrightarrow{L} \begin{bmatrix} r_w & \frac{1}{2}\tilde{\alpha} r_w (r_1 + r_2) \\ \frac{1}{2}\tilde{\alpha} r_w (r_1 + r_2) & \frac{1}{3}\tilde{\alpha}^2 (r_2^3 - r_1^3) \end{bmatrix} = \mathbf{V} \tag{C.2}$$

and the matrix $\Upsilon_T^{-1} \left[\sum_{t=[Tr_1]}^{[Tr_2]} X_t \varepsilon_t \right]_{(-2)\times 1}$,

$$\begin{bmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \varepsilon_t \\ \Sigma y_{t-1} \varepsilon_t \end{bmatrix} = \begin{bmatrix} T^{-1/2}\Sigma \varepsilon_t \\ T^{-3/2}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{p} \begin{bmatrix} T^{-1/2}\Sigma \varepsilon_t \\ T^{-3/2}\Sigma \tilde{\alpha}(t-1)\varepsilon_t \end{bmatrix}$$

The variance-covariance matrix of the last two elements,

$$E\begin{bmatrix} T^{-1/2}\Sigma\varepsilon_t \\ T^{-3/2}\Sigma\tilde{\alpha}\left(t-1\right)\varepsilon_t \end{bmatrix}\begin{bmatrix} T^{-1/2}\Sigma\varepsilon_t \; T^{-3/2}\Sigma\tilde{\alpha}\left(t-1\right)\varepsilon_t \end{bmatrix} = \sigma^2 W,$$

where

$$W = \begin{bmatrix} 1 & \frac{1}{2}\tilde{\alpha}\left(r_2^2 - r_1^2\right) \\ \frac{1}{2}\tilde{\alpha}\left(r_2^2 - r_1^2\right) & \frac{1}{3}\tilde{\alpha}^2\left(r_2^3 - r_1^3\right) \end{bmatrix}.$$

Therefore, we have

$$\begin{bmatrix} T^{-1/2}\Sigma\varepsilon_t \\ T^{-3/2}\Sigma y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{L} h_2 \sim N\left(\mathbf{0}, \sigma^2\mathbf{W}\right) \tag{C.3}$$

Combining equation C.2 and equation C.3, it follows that

$$\begin{bmatrix} T^{1/2}\left(\hat{\alpha} - \alpha\right) \\ T^{3/2}\hat{\beta} \end{bmatrix} \xrightarrow{L} \mathbf{V}^{-1}h_2 \sim N\left(\mathbf{0}, \sigma^2\mathbf{V}^{-1}W\mathbf{V}'^{-1}\right).$$

Let $R_1 = r_1 + r_2, R_2 = r_2^2 - r_1^2 = r_w R_1, R_3 = r_2^3 - r_1^3$;

$$\begin{aligned} \mathbf{V}^{-1}W\mathbf{V}'^{-1} &= \begin{bmatrix} r_w & \frac{1}{2}\tilde{\alpha}r_w R_1 \\ \frac{1}{2}\tilde{\alpha}r_w R_1 & \frac{1}{3}\tilde{\alpha}^2 R_3 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \frac{1}{2}\tilde{\alpha}rR_1 \\ \frac{1}{2}\tilde{\alpha}rR_1 & \frac{1}{3}\tilde{\alpha}^2 R_3 \end{bmatrix} \begin{bmatrix} r_w & \frac{1}{2}\tilde{\alpha}r_w R_1 \\ \frac{1}{2}\tilde{\alpha}r_w R_1 & \frac{1}{3}\tilde{\alpha}^2 R_3 \end{bmatrix}^{-1} \\ &= \frac{1}{16R_3^2 - 24r_w^2 R_1^3 + 9r_w^2 R_1^4} \\ &\begin{bmatrix} 4R_3\left(-3r_w^2 R_1^2 + 4R_3\right)/r_w^2 & -6R_1\left(-3r_w^2 R_1^2 + 4R_3\right)/r_w\tilde{a} \\ -6R_1\left(-3r_w^2 R_1^2 + 4R_3\right)/r_w\tilde{a} & 12\left[4R_3 + 3\left(1 - 2r_w\right)R_1^2\right]/\tilde{a}^2 \end{bmatrix} \end{aligned}$$

We can see that $\hat{\beta}$ converges at rate $T^{3/2}$ to a Gaussian variable

$$T^{3/2}\hat{\beta}.\psi\left(1\right) \xrightarrow{L} h_3 \sim N\left(\mathbf{0}, \sigma^2\mathbf{\Psi}\right),$$

where

$$\mathbf{\Psi} = \frac{12\left[4R_3 + 3\left(1 - 2r_w\right)R_1^2\right]}{\left(16R_3^2 - 24r_w^2 R_1^3 + 9r_w^2 R_1^4\right)\alpha^2}.$$

For the sup ADF test, $R_1 = r_w, R_3 = r_w^3$,

$$\mathbf{\Psi} = \frac{12\left[4r_w + 3\left(1 - 2r_w\right)\right]}{r_w^3\left(25r_w - 24\right)\alpha^2}.$$

For the ADF test, $R_1 = 1, R_3 = 1$; $\mathbf{\Psi}$ $=12/\alpha^2$.Therefore, all three types of the ADF statistics can be compared with the usual t for an asymptotically valid test.